UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED
ST/ECA/SM/11
9 September 1981
Original : ENGLISH

ECONOMIC COMMISSION FOR AFRICA
Seminar on Household Surveys
Addis Ababa, 21-30 September 1981

LC/NAIROBI

ON SURVEY DATA ANALYSIS

CONTENTS

	Paragraphs	Pages
I. INTRODUCTION	1 - 6	1
II. SOME POLICY QUESTIONS	7 - 17	1 - 2
III. THE EXPLORATION OF DATA STRUCTURES AND MODEL FITTING	18 - 27	3 - 8
General	18 - 22	3
Analysis of Contingency Tables	23	3 - 4
Cluster Analysis	24	4 - 5
Aid Analysis	25	5 - 6
Box-Jenkins Analysis	26	6
Model Fitting	27	6 - 8
IV. BIBLIOGRAPHY	-	9 - 11

I. INTRODUCTION

1. A paper on survey data evaluation and analysis (E/CN.14/SM/27) was produced for the second meeting of the working group on organization, content and methodology of household surveys in October 1979. In the last meeting of the working group (June-July 1981) another paper on survey data analysis was presented (E/CN.14/SM/39).

2. The last paper dealt with analytical requirements and responsibilities, activities prior to analysis, analytical options, the data base approach and some notes on analysis. Rather little was said about the analytical tools for analyzing survey data.

3. The main purpose of the present paper is to give a brief account of a series of techniques for the exploration of data structure and analysis of survey data. A bibliography of relevant books is included. Completeness has not been attempted. One reason is that there is no particular limit to the scope and applicability of statistical theory and method in this connection.

4. The working group has earlier strongly stressed the importance of a close cooperation between statisticians and users of statistics in order to be able to formulate the statistical problems correctly and to present and analyse the data so as to be useful for decision making and action taking. This topic is no doubt of utmost importance and should be dealt with in some detail when discussing the AHSCP.

5. While recognizing the importance of proper problem formulation, it is believed that there is a strong connection between the success achieved in this area and knowledge of statistical tools available for describing and analyzing survey data. There is a two-way process.

6. A prerequisite for applying techniques of exploring data structures and analysis of data is that the previous stages of survey work has been well handled. It means that added importance is attached to an adequate sampling design, control of non-sampling errors, etc.

II. SOME POLICY QUESTIONS

7. The question if producers of statistics should carry out consumer-oriented analysis of official statistics has been much discussed in recent years. This theme was for example a topic of discussion at the ISI-session in Warszawa in 1975. The question has also been discussed at several meetings of the UN economic commissions.

8. It has been generally agreed that the producers of statistics should carry out producer-oriented analysis. This includes analysis of data quality, calculations of sampling errors, etc. The purpose with this type of analysis is to reveal limitations concerning the data collected and to warn the users from drawing wrong conclusions. When it comes to consumer-oriented analysis - which is the subject of this paper - different opinions have been expressed. However, the current trend is that the users of statistics request more elaborate statistical products as a basis for decision-making.

9. It is believed that a statistical office is the proper place for the presentation and analysis of survey data, provided that there is a close cooperation between statisticians and planners at various stages of the survey work. One obvious reason is the statistical nature of the techniques applied as well as

possession of first hand knowledge of the quality of the data collected. Of course, this should not prevent users of data to get engaged in analytical work as well.

10. The duty of the statistical office to carry out consumer-oriented analysis has a series of consequences both with regard to statistical training and to statistical organization. There are also budgetary consequences, as more resources have to be allocated to the planning and analysis of surveys.

11. The importance of training can hardly be exaggerated. Multivariate analysis is usually not extensively dealt with at lower academic levels. Accordingly, post-graduate courses have to be organized to meet the needs of education in this discipline of statistics.

12. In addition to university training it appears necessary to organize in-service training in the central statistical offices. This in turn calls e.g. for strenghtening of the statistical libraries, which should have an adequate supply of textbooks and journals in the field of statistical methodology.

13. For each survey there should be a project officer, who is involved at all stages, from the start and up to the writing of reports. His first duty should be to work closely with planners and users of the survey results to make the reports consumer-oriented. He will have to be backed up at various stages by specialists in fields such as administration, organization, sampling, data processing, statistical analysis, etc. However, he should himself have a broad general knowledge of various areas, particularly methods of statistical analysis.

14. It is advantageous to set up a section, consisting of mathematically oriented statisticians who are specialists in fields such as sampling and analysis. Their task should be to assist the project officers, to study continuously new statistical methodology and to arrange in-service training for other statisticians in the office.

15. Concerning the budgetary aspects it should be kept in mind that the additional cost of more advanced description and analysis of household surveys is small in comparison with the other costs of interviewing, transport, checking, coding etc. The marginal utility is high. No doubt statistics will be more highly appreciated as a consequence of analytical work .

16. Access to computers is fundamental in the analysis of surveys. There are a great many computer programs for the purpose. Attemps should be made to build up useful program libraries in the statistical offices. It is suggested that the ECA makes a contribution in this respect.

17. It would be valuable if an African Statistical Journal could be set up with support from the ECA, containing e.g. articles on applications of statistical analysis of survey data in the African context. This would certainly promote the African household survey programme.

III. THE EXPLORATION OF DATA STRUCTURES AND MODEL FITTING

18. Until recently, the almost only type of sample survey analysis consisted of estimations of population means, totals, frequency distributions, maybe sampling eerors, correlation coefficients, cross-tabulations, graphs and the like. If well done and carefully examined, much information can be obtained from such exercices, also with regard to relationships between variables.

19. Particularly two factors have transformed the analysis of survey data in the last decade: the development of new techniques for data analysis and the availability of powerful computers to survey analysts.

20. Until rather recently statistical methods for dealing with relationships between variables have been developed to meet the needs of the experimental scientist. The social and economic research workers on the other hand, have mostly to use surveys, which result in a complex and often large set of data. There is a need for exploratory techniques, which adequately describe structural features in the data set. This may form a prelude to theory generation and model construction.

21. Many writers on the subject distinguish between techniques which aim at a more refined form of data description (exploratory techniques) on the one hand and techniques which aim at showing causal relationships (model fitting) on the other hand. The distinction may sometimes be a bit difficult to make. Anyway, in view of the relative scarcity of survey data in the African region the techniques indicated in the following mainly belong to the exploratory techniques. Only one paragraph deals with model fitting explicitly.

22. The choice of statistical techniques, which are briefly presented here, is selective. The same goes for the bibliography given. The aim of the paper is rather to draw attention to the importance of considering modern methods of analysis when discussing the African household survey programme. Practical applications will certainly give interesting experiences and add more substance to the programme.

23. Analysis of contingency tables. This type of analysis is mainly concerned with methods of dealing with frequency data occuring in the form of cross-classifications or contingency tables, such as 2x2 contingency tables, rxc contingency tables or multidimensional tables. When a population is classified into several categories one can count the individuals in each category. These counts or frequencies are the type of data with which this kind of analysis is primarily concerned. The analysis is dealing with qualitative rather that quantitative data obtained from measurement of continuous variables. Of course, continuous data can often be put into discrete form by the use of intervals on a scale.

It is to a large extent a matter of hypothesis testing techniques for the analysis of contingency tables. Among these may be mentioned the chi-square test, Fisher's exact test for 2x2 tables, Mc Nemar's test for correlated proportions in a 2x2 table, Gart's test for order effect, the $V(\bar{x}^2)$ method when combining information from several 2x2 tables, Cochran's method for a similar stituation, Lancaster and Irwin's method for partitioning rxc tables, etc.

Frequently one may be interested in measuring the strength of association between the qualitative variables involved in addition to testing the significance. A number of measures have been constructed, such as the mean square contingency coefficient $\phi^2 = x^2/N$, Pearson's acoefficient of contingency

$P = \sqrt{\left(\frac{x^2/N}{1+x^2/N}\right)}$, Kendall's and Stuart's $T = \frac{x^2/N}{\sqrt{(r-1)(c-1)}}$

Cramer's $C = \frac{x^2/N}{\min(r-1,c-1)}$, Goodman and Kruskal's lambda and Gamma measures and Somer's d-measure.

Another approach has also been considered, namely that of fitting models and of estimating the parameters in the models. The starting-point is a theory about the observations and the parameters in the model represent the effects that particular variables or combinations of variables have in determining the values. Of particular interest are methods developed for fitting log-linear models to multidimensional frequencey data and the description of these data in parametric terms. Among statisticians who have been engaged in developing the methods in question may be mentioned Bartlett, Bock, Haberman, Newton, Raphson, Deming, Stepham, Birch, Bishop, Fienberg and Goodman. The most comprehensive account is given in Bishop, Fienberg and Holland (1975) : Discrete Multivariate Analysis. Another book dealing with the subject is Everitt (1977): The Analysis of Contingency Tables. Both volumes are included in the bibliography given.Finally it may be mentioned that a number of computer programs for the fitting of log-linear models to complex multi-dimensional contingency tables are available.

24. Cluster analysis. This is a label for a number of different numerical methods of classification. A common feature is that these methods - based on some criterion of similarity - group variables or objects in clusters. The units in a certain cluster should be as homogenous as possible while the clusters between each other should be as different as possible.

Several measures of similarity are described in e.g. Sneath and Sokal : Principles of Numerical Taxonomy. The most common measures are the Euclidean distance and the correlation coefficient. Measures have been defined for quantitative as well as qualitative variables.

Hierarchial techniques are probably the most commenly used class of clustering methods. They may be subdivided into agglomerative methods which proceed by a series of successive fusions of individuals into groups and divisive methods which partition a set of n individuals successively into finer partitions. The results of both agglomerative and divisive techniques may be presented in the form of a s.c. dendogram, which is a two-dimensional diagram illustrating the fusions or partitions which have been made at each successive level.

Another class of clustering techniques are those which seek to partition the set of data so as to optimize some predefined clustering criterion. These techniques assume that the number of groups into which to partition the data has been set by the user to some value, say k. Many clustering criteria have been suggested. Probably the most common is minimization of the pooled within cluster sum of squares, a method proposed by e.g. Singleton and Kantz, Mac Queen and Ball and Hall.

Cluster analysis can be utilized for descriptive purposes by meaningful grouping of a large number of individual data. It can also be used to explore the data structure to obtain a basis for model building or for hypothesis generating purposes. When evaluating the results of cluster analysis it is important to make use of available knowledge of the subject-matter.

A common problem when applying cluster analysis is that the variables used may have been measured by different types of scales. Then one has to decide upon some specific type of scale and then transform all the variables into that scale. Of course, this leads to a certain loss of information.

The number of papers describing applications of clustering techniques is considerable. There have been several reported applications in psychiatry, psychology, education and market research. Cluster analysis has been used for establishing a nomenclature for occupations and for studying change of occupations. It has also been used to study areas within cities. Finally it may be mentioned that cluster analysis has been used to form a basis for sampling design, e.g. as a stratification tool using as a criterion of effectiveness the ability to increase the efficiency of estimation.

There are many packages of computer programs available for various methods of cluster analysis.

In addition to the volume by Sneath and Sokal may be mentioned Ball: Classification Analysis, Everitt: Cluster Analysis, Jardine and Sibson: Mathematical Taxonomy, Singleton and Kantz: Minimum Square Error Clustering Algoritm and various publications by Wolfe, e.g. NORMIX: Computational Methods for estimating the parameters of Multivariate Normal Mixtures of Distributions. Reference is made to the bibliography attached.

25. AID Analysis. AID means Automatic Interaction Detection, which indicates that the method is designed to detect interactions in a data set. AID was introduced in 1964 by Morgan and Sonqvist at the University of Michigan. They also made ADP-programs. As an analytic method AID has developed along with the program, which is characteristic of other techniques of exploring data structures, such as cluster analysis. It is usually impossible to manage the calculations without the use of dator capacity.

A group of n individuals is under study, each characterized by a set of variables $(y, y_2 \ldots y_m; x_1, x_2, \ldots x_p)$. The y_i are taken to be the dependent variables, x_j independent or predictor variables. In some contexts there is just one dependent variable y. AID is a technique of step by step binary segmentation. Firstly, the sample as a whole is split into two subgroups and the same procedure is applied sequentially to resulting subgroups. However, at any stage decisions must be made on the particular predictor variable to be used in inducing the split, and how the two subsets of its values are to be formed. Since the objective is to explain the variation in the dependent variables, it seems sensible to make these two decisions so that the resulting subgroups are as different as possible and are internally as homogeneous as possible with regard to the y_i's.

The basic strategy is to examine each predictor variable x_j and to choose a partition of its set of values such that the induced binary split results in subgroups that maximize the chosen "distance" function. The next step is to choose among these the variable which produces the overall maximum. In a sense it is the best explanation of the y_i one can get through a single split into two groups. The process continues by sequentially splitting the subgroups obtained from the previous stages using the same criterion for deciding which predictor to use. Rules have been given for deciding when to stop the process.

The result of the binary segmentation process is a hierarchical tree. This is a diagram, which gives information on the terminal or final groups, association between predictors and interaction effects.

The technique has been widely applied, e.g. for studies of salaries, prices, rents, effects of television broadcast, smoking, problem drivers, emigration, education etc. It has also been used as a tool for planning sample surveys.

Particular reference is made to Sonquist : Multivariate Model Building. The Validation of a Search Strategy, Sonquist, Baker and Morgan : Searching for Structure and Morgan Messenger : THAID. A Sequential Analysis Program for the Analysis of Nominal Seale Dependent Variables.

26. Box-Jenkins Analysis. This technique refers to a special strategy analysis. Where the main purpose is to give short forecasts of time series. The method can more generally be used for description and analysis of time series. One might for instance test some hypothesis concerning changes of trends as a consequence of external events, such as strikes, new regulations or energy crises.

The Box-Jenkings method calls for extensive numerical calculations and it is therefore necessary to have access to dator capacity and special ADP-programs to carry out the analysis.

Box-Jenkins analysis can be carried out at two levels. The first is the level of univariate models, when the analysis is based only on the historical development of specific series. The results are used to eliminate seasonal effects and to identify points, where the quality of data may be defective. The second level is that of building up transfer function models taking into account dependent series as well as a number of explanatory series. Transfer function models may be used for a) short forecasts b) improvement of preliminary statistics with the help of leading indicators c) estimation of the effects of internal/external events and perhaps making corrections for these effect.

The method has been utilized in many contexts, e.g. improvement of data quality, improvement of preliminary indices, studies of the effect of the oil crises on the consumption of electricity by the households, etc.

Reference is particularly made to Box and Jenkins :"Time Series Analysis" and Jenkins: "Practical Experiences with Modelling and Forecasting Time Series".

27. Model Fitting. This term relates to the situation when one has reached an insight into the causal relationships between variables in some subject area. Thus a model of the form $(y_1, y_2, \ldots y_m) = f(x_1, x_2 \ldots x_p) + \epsilon$ can be set up where the functional relationships are known. This make it possible not only to fit a model, but also to test the goodness of fit of the model to the observed data.

Of course a model is only an approximation to the real world. It is usually impossible to include all variables relavant to the problem dealt with, which is reflected in the size of the error term ε.

The type of models most commonly used in survey analysis is the class of linear models. A dependent variable is assumed to be a linear function of the parameters of the model. The relationship is $f(y) = \sum_{j=1}^{p} \beta_j x_j + \varepsilon$ where $f(y)$ is a dependent variable, β_j's are the parameters and the x_j's are the explanatory variables. Interactions can be built into the model by defining new variables which represent these interactions. This is a problem which has led to the development of search techniques, some of which have been indicated earlier in this paper.

The most commonly used form of the general linear model is ordinary least squares regression. Additional assumptions must be made about the distribution of the ε's if the parameters are to be estimated optimally. The disturbances must be uncorrelated with each other and with the x's and must all have equal variance. The analysis of variance is a special case of multiple regression which arises when all the explanatory variables are nominal-scale or categorical variables, while the dependent variable is still a continuous interval-scale variable. The analysis of covariance may be used when some of the explanatory variables are continuous and some are nominal.

Frequently one has to assume several variables which have to be considered simultaneously. Two ways of dealing with such situations are multiple regression and simultaneous equation estimation. The difference between these two methods is that in the latter case there will be more than one dependent variable. For simultaneous equation estimation a variable may appear as dependent in one equation and as an explanatory variable in others. This technique has been extensively used in the field of econometrics.

Canonical correlation is another technique that represents an extension of linear-regression analysis. It is concerned with two sets of variables. The method comprises the derivation of a variate - a linear combination of the variables in a set - for each set such that the correlation between the sets are maximized. To date, applications of the method have been principally in economics for the estimation of demand and supply relationships, where both price and quantity may be expressed as a linear combination of variables.

The basic hypothesis of discriminant analysis is that a population is assumed to be made up of two distinct subpopulations. It is further assumed that it is possible to find a linear function of certain measures or attributes of the population that will allow an observer to discriminate between the two subpopulations. Discriminant analysis - particularly developed by Fisher, Neyman and Pearson - was originally used in the biological sciences. However, the technique has many possible applications in the social sciences. It has for instance been occasionally used in the field of transport statistics.

Factor analysis comprises a set of widely differing techniques which have been developed since Spearman introduced the conceptual framework in the beginning of the 20th century. Principal component analysis, on the other hand , is a more well-defined procedure. Basically these methods attempt to condense a set of variables, that is, the factors or components, which have not been directly measured The methods have their roots in the basic principles of multivariate linear-regression analysis and there has been some controversies as to the nature of the methods. Kendall has suggested that principal component analysis is the search through data to try to find components that may reduce the dimensions of variation while

factor analysis starts with the hypothesis of a model and tests it against the data to see if it fits. Factor analysis was originally developed for use in the field of psychology but has later on been applied in various other fields. It is not sufficient to obtain just a mathematical solution. The solution should also be conceptually interpretable. Frequently this is one of the major difficulties encountered in factor analysis.

Two other methods of model-fitting may be briefly mentioned. Path analysis has become one of the standard tools for analysing interval-level data for which an explicit causal structure can be stipulated. Probit-and logit analysis represent techniques to be used in a situation when one wishes to estimate the probability of some occurence as a function of a set of independent variables that may take values over a wide range.

There is a vast literature on the various model-fitting techniques.

IV. BIBLIOGRAPHY

Anderberg, M.R. (1973) ; Cluster Analysis for Applications (Head Press, New York)

Andersson, T.W. (1958) : An Introduction to Multivariate Statistical Analysis (Wiley, New York)

Andrews, F .M and Messenger, R.C. (1973) : Multivariate Nominal Scale Analysis (Inst. for Social Research, Univ. of Michigan)

Armitage, P. (1971) : Statistical Methods in Medical Research (Blackwell Scient. Publ., Oxford, Edinburgh)

Ball, G.H. (1971) : Classification Analysis (Stanford Research Institute SRI Project 5533)

Beale, E.M.L. (1969): Cluster Analysis (Scientific Control Systems, London)

Bibby, J. and Toutenburg, H.(1977) : Prediction and Improved Estimation in Linear Models (Wiley, London)

Bishop, Y.M., Fienberg, S.E. and Holland, P.W. (1975): Discrete Multivariate Analysis; Theory and Practice (The M.I.T. Press, Cambridge, Massachusetts)

Borgatta, E.F. (1969) : Sociological Methodology 1969 (San Francisco : Jossey-Bass)

Box, G.E.P. and Jenkins, G.M. (1976) : Time Series Analysis. Forecasting and control. (Holden Day)

Cassel, C-M., Sarndal, C-E. and Wretman, J.H. (1977) : Foundations of Inference in Survey Sampling (Wiley, New York)

Cattell, R.B. (1952) : Factor Analysis (Harper, New York)

Cox, D.R. (1970): Analysis of Binary Data, (Methuen, London)

Cramér, H. (1946) : Mathematical Methods for Statistics (Princetown Univ. Press)

Davis, D. : A factorial ecology of Malmö 1965; the social geography of a city (Lunds Universitets Geografiska Institution, Lund, Sweden)

Davis, J.A. (1971) : Elementary Survey Analysis (Englewood Cliffs, N.J. : Prentiee Hall)

Dhrymes, P.J. (1970) : Econometrics : Statistical Foundations and Applications (Harper and Row, New York)

Draper, N.R. and Smith, H. (1966) : Applied Regression Analysis (Wiley, New York)

Everitt, B.S. (1974) : Cluster Analysis (Heinemann Press, London)

Everitt, B.S. (1977) ; The Analysis of Contingency Tables (Chapman and Hall, London)

Finney, D.J. (1971) : Probit Analysis (Univ. Press, Cambridge)

Fisher, R.A. (1950) : Statistical Methods for Research Workers (Oliver and Boyd, Edinburgh)

Fleiss, J.L. (1973) : Statistical Methods for Rates and Proportions (Wiley, New York)

Frankel, M.R. (1971) : Inference from survey samples (Inst. for social research, Univ. of Michigan)

Godambe, V.P. and Sprott, D.A. (1971) : Foundations of Statistical Inference (Symposium on the Foundations of Statistical Inference held at the Department of Statistics, Univ. of Waterloo, Canada)

Goldberger, A.S. (1968) : Topics in Regression Analysis (Collier-Macmillan, London)

Goldberger, A.S. and Duncan, O.D. (1973) : Structural Equations Models in the Social Sciences (Seminar Press, New York)

Gorsuch, R.L. (1974) : Factor Analysis (W.B. Saunders Co, Philadelphia)

Haberman, S.J. (1974) : The Analysis of Frequency Data (Univ. of Chicago Press, Chicago)

Hags, W.L. (1973) : Statistics for Social Sciences (Holt, Reinhort and Winston, New York)

Hanushek, E.A. and Jackson, J.E. (1977) : Statistical Methods for Social Scientists (Academic Press, New York)

Harman, H.H. (1960) : Modern Factor Analysis (Univ. of Chicago Press, Chicago)

Hollander, M. and Wolfe, D.A. (1973) : Non-Parametric Statistical Methods (Wiley, New York)

Hope, K. (1968) : Methods of Multivariate Analysis (Univ. of London Press, London)

Jardine, N. and Sibson, R. (1971) : Mathematical Taxonomy (Wiley, London)

Jenkins, G.M. (1979) : Practical experiences with Modelling and Forecasting Time Series Jenkins and Partners Ltd)

Johnston, J. (1972) : Econometric Methods (Mc Graw-Hill, New York)

Kendall, M.G. (1965) : A course in Multivariate Analysis (Griffin, London)

Kendall, M.G. and Stuart, A. (1961) : The Advanced Theory of Statistics (Griffin, London)

Kendall, M.G. (1955) : Rank Correlation Methods (Hafner, New York)

Kirk, R.E. (1968) : Experimental Design Procedures for the Behavioural Sciences (Brooks & Cole, Calif)

Kish, L. (1965) : Survey Sampling (Wiley, New York)

Lawley, D.N. and Maxwell, A.E. (1971) : Factor Analysis as a Statistical Method (Butterworths; London)

Leser, C.E.V. (1974) : Econometric Techniques and Problems (Griffin, London)

Lindsey, J.K. (1973) : Inference from Sociological Survey Data (Elsevier, Amsterdam)

Mac Naughton-Smith, P. (1965) : Some Statistical and Other Numerical Techniques for Classifying Individuals (Home Office Research Unit Report No. 6, London)

Miller, R.G. Jr. (1966): Simultaneous Statistical Inference (Mc Graw-Hill, New York)

Morgan, J.N. and Messenger, R.C. (1973) : THAID, A Sequential Analysis Program for the Analysis of Nominal Scale Dependent Variables (Inst. for Social Research, Univ. of Michigan)

Moser, C.A. and Kalton, G. (1971) : Survey Methods in Social Investigations (Heinemann, London)

Mosteller, F. and Tukey, J.W. (1977) : Data Analysis and Regression (Addison-Wesley, London etc.)

Namboodiri, N.K. Carter, L.F. and Blalock, H.M. (1975) : Applied Multivariate Analysis and Experimental Designs (McGraw-Hill, New York)

Neter, J.A. and Wasserman, W. (1974) : Applied Linear Statistical Models (Irvin, Homewood, Illinois)

O'Muircheartaigh, C. and Payne, C. (1977) : Exploring Data Structures (Wiley, London)

O'Muircheartaigh, C.A. and Payne, C. (1977) : Model Fitting (Wiley, London)

Pielov, E.C. (1969) : An Introduction to Mathematical Ecology (Wiley, New York)

Plackett, R.L. (1974) : The Analysis of Categorical Data (Griffin, London)

Rao, C.R. (1952) : Advanced Statistical Methods in Biometric Research (Wiley, New York)

Rao,C.R. (1968) : Linear Statistical Inference (Wiley, London)

Rao, P. and Miller, R.L. (1971) : Applied Econometrics (Wadsworth Publ. Co, Belmont)

Rummel, R.J. (1973) : Applied Factor Analysis (Evanston, III; Northwestern Univ. Press)

Saker, G.A.F. (1977): Linear Regression Analysis (Wiley, New York)

Searle, S,R. (1971) : Linear Models (Wiley, New York)

Schuessler, K. (1971) : Analyzing Social Data: A Statistical Orientation (Houghton Mifflin, Boston)

Singleton, R.C. and Kautz, W. (1965) : Minimum Square Error Clustering Algoritm (Res. Inst., Stanford)

Sneath, P.H.A. and Sokal, R.R. (1973) : Principles of Numerical Taxonomy (Freeman, London)

Sonquist, J.A. (1970) : Multivariate Model Building (Inst. for Social Research, Univ. of Michigan)

Sonquist, J.A. Baker, E.L. and Morgan, J.N. (1971) :Searching for Structure (Inst. for Social Research, Univ. of Michigan)

Sonquist, J.A. and Morgan, J.N. (1974) : The Detection of Interaction Effects (Ann Arbor, Inst. for Social Research)

Sprent, P. (1969) : Models in Regression and Related Topics (Methuen, London)

Stopher, P.R. and Meyburg, A.H. (1979) : Survey Sampling and Multivariate Analysis for Social Scientists and Engineers (Lexington Books, Massachusetts)

Theil, H (1973) : Principles of Econometrics (Wiley, London)

Thurstone, L.L. (1947) : Multiple Factor Analysis (Univ. of Chicago Press, Chicago)

Tintner, G. (1952) : Econometries (Wiley, New York)

Tukey, J.W. (1977) : Exploratory Data Analysis (Addison-Wesley, London etc.)

Van de Geer, J.P. (1971): Introduction to Multivariate Analysis for the Social Sciences (Freeman, San Francisco)

Williams, E.J. (1959) Regression Analysis (Wiley, New York)

Wolfe, J.H. (1967) : NORMIX : Computational Methods for estimating the parameters of Multivariate Normal Mixtures of Distributions (Res. Memo SRM 68-2, San Diego Navy Personnel Research Activity)

Wolfe, J.H. (1969) : Pattern Clustering by Multivariate Mixture Analysis (Res. Memo SRM 69-19, San Diego: US Navy Personnel Research Activity)

Tyron, R.C. and Bailey, D.E. (1970) : Cluster Analysis (Mac Graw-Hill, New York)

Yates, F.S. (1981: Fourth Edition) Sampling Methods for Censuses and Surveys (Griffin, London)

UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED
ST/ECA/SM/12
26 August 1981
Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA
Seminar on Household Surveys
Addis Ababa, 21-30 September 1981

SOME ASPECTS OF SOCIAL SURVEYS*

C O N T E N T S

		Paragraph	Page
I.	SOCIAL SURVEYS AS A BASIS FOR THE DEVELOPMENT OF OBJECTIVE AND SUBJECTIVE SOCIAL INDICATORS . .	1 - 4	1
II.	MAIN TOPICS OF SOCIAL SURVEYS	5 - 7	1 - 2
III.	SURVEY APPROACH TO DIFFERENT TOPICS	8 - 13	2 - 3
IV.	THE GERMAN HEALTH SURVEY	14 - 26	3 - 6
V.	THE GERMAN HOUSING SAMPLE SURVEYS	27 - 37	6 - 8
VI.	COMPARABILITY OF SOCIAL SURVEY DATA	38 - 39	9

ANNEXES

ANNEX I COMPREHENSIVE SURVEY (OECD): QUESTIONNAIRE ON HEALTH
ANNEX II FOLLOW-UP SURVEY 'SICKNESSES AND ACCIDENTS' (FRG)
ANNEX III HOUSING SAMPLE SURVEY: QUESTIONNAIRE (FRG)
ANNEX IV THE 1980 OECD LIST OF SOCIAL INDICATORS

* This paper was prepared by Mr. H.L. Mayer, Federal Statistical Office, Federal Republic of Germany. It is presented in its original English version without editing. The views expressed in it are those of the author and do not necessarily reflect those of the United Nations.

LC/NAIROBI

I. SOCIAL SURVEYS AS A BASIS FOR THE DEVELOPMENT OF OBJECTIVE AND SUBJECTIVE SOCIAL INDICATORS*

1. Over the last ten years or so a new data source directly designed for the development of social indicators has emerged in some countries, namely the multi-purpose social survey covering the whole - (compared with the Microcensus) or a large part - of the social field. The interest of these surveys lies in the following considerations:

2. Specialized surveys or administrative data sources are frequently designed to meet the needs of specific users. These needs are often very far removed from social indicators. The inadequacy of these sources lies in the fact that they are primarily intended for the preparation of sectoral programmes and policies. For example, in the surveys on housing, it is often the dwelling itself which is the main focus, whereas in the perspective of social indicators on housing the unit of analysis will not be the actual dwelling but the social unit using it, i.e. the individual, family, or household. In short, since the objectives of specialized surveys or administrative sources are not the measurement of individual well-being, operational definitions, population coverage, measuring units, etc., are often inadequate for social indicators.

3. Specialized surveys and other sources do not permit any general analysis of levels and distribution of individual well-being by linking up all these various aspects. On the other hand, a multi-purpose social survey can be a relatively economical and effective means of attaining this general analysis from a normative viewpoint. It may also contribute to an explanatory analysis on specific points (although to a relatively limited extent, as will be seen below). But obviously such a type of survey cannot provide information concerning the evaluation of specific sectoral programmes or policies.

4. In this context social surveys should make it possible:

- to measure the [illegible] of individual well-being in each of the broad social areas included in the survey;
- to establish a number of normative and possibly explanatory links between the social situations existing in the various areas.

II. MAIN SECTORS OF SOCIAL SURVEYS

5. Most of the social surveys in European countries seek to obtain exhaustive information on the following basic social areas:

a) health and use of health services
b) housing conditions
c) employment and working conditions
d) education.

---

*See: 'Inventory of data sources for social indicators' (Organization for Economic Co-operation and Development).

6. Other areas may be covered partially and integrated into the sections of the survey concerning one or other of the four basic areas. This is generally the case for environmental conditions which are covered in the sections of the questionnaire concerning housing conditions and working conditions, and which are approached only from the perceptive angle (perception by respondents of annoyances relating to the housing environment and working conditions). Some areas are covered only to obtain reference data, in order to establish the bases for specific disaggregations. The most frequent case is wealth and income.

7. Lastly, other centres of interest may be covered more or less broadly in certain surveys. They are essentially those concerned with the analysis of social mobility and intergenerational changes in living conditions. In the questionnaires, apart from the standard questions on the composition of the basic social unit (family, household), we may also find questions concerning the individual characteristics of the other members of the unit and those of their parents and children (these questions deal mainly with health status, educational level and employment).

III. SURVEY APPROACH TO DIFFERENT TOPICS

8. Two types of approach may be used jointly in the social surveys:

- The "objective" approach: in this case the aim is to obtain information on the real concrete situation of the respondent with regard to the area in question. This approach may take two forms: either descriptive (the respondent is questioned about the exact quantitative characteristics of the items considered) or perceptive (the respondent must judge himself whether the items in question have this or that characteristics for which no exact quantitative criteria are given in the questionnaire). For instance, information about quality of teeth (health area) may be requested in two complementary ways: by asking the respondent how many times he visited the dentist in the year preceding the survey (objective data) and by asking him whether he can chew a given food without difficulty (perceptive approach).
- The "subjective" approach: in this case the aim is to obtain information on the respondent's personal opinion regarding this or that item of the area in question. For instance, he may be asked to give his opinion about work safety in his form by picking an appropriate rating from a list ranging from very good to very bad.

9. Obviously it is often difficult to assess what type of approach corresponds to particular sections of the survey, whether the choice is between descriptive or perceptive, or between perceptive and subjective. The choice of an approach for each item included in the survey depends on:

- the orientation given to the study of this item in the survey;
- the importance assigned to this item in the survey;
- the specific characteristics of this item.

10. Thus for instance it is impossible to determine the health status of the population in a purely descriptive manner on the basis of multi-purpose social surveys. Accordingly such surveys merely include two types of questions (see Annex 1):

- Firstly, questions regarding the use of health services (usable only for serious illness, accident or impairment, for in the event of non-serious illness, accident or impairment the use of health services depends mainly on social habits and the efficiency of the services).
- Secondly, indirect information on the respondent's health (questions relating to the respondent's perception of this health, generally by referring to such characteristics as ability or inability to run or walk a given distance, to dress unaided, etc.).

11. To take another example, the form of social surveys (limited duration and inability to establish extensive control methods) limits the extent of the study of the conditions of the physical environment at home and at work to perceptive questions on environmental characteristics (e.g., as regards noise, the respondent will be asked: during the night/day is the housing environment: followed by a set of possible responses, graded from "very noisy" to "very quiet"; see Annex III: German Housing Sample Survey; and Annex IV: The 1980 OECD List of Social Indicators).

12. The general orientation of the survey largely influences the type of approach used for each subject area. For instance, the German Household Survey (Microcensus, questions on health; see Annex II) makes practically no use of the subjective approach and therefore use only a descriptive-perceptive approach; but this is not the case for other surveys which for a large number of areas make ample use of questions on level of satisfaction and subjective appraisal.

13. Lastly, the type of approach used in the questionnaires for each subject area depends on the importance and role assigned to that area. If an area is designed only as a means of collecting reference data, descriptive questions will very often be scarcely developed and use will be made instead of far less precise perceptive questions or questions approaching the subject indirectly. This applies for instance to the topic of income and wealth. It is recognized that in general social surveys it is impossible to construct a proper questionnaire in this area as the information obtained is in any case of extremely doubtful validity. Consequently questions are generally confined to a few on sources of income, followed by one or two on level of money income, and are either open-ended ("What was your income for the month/year...?" expressed in monetary units) or more often closed ("In what range did your income for the month/year ... fall?") with a choice between various income brackets).

IV. THE GERMAN HEALTH SURVEY

14. In the very beginning when the work on Health Surveys was started in the Federal Republic of Germany, a voluntary pilot survey on illness and accidents was for the first time conducted at the federal level. Its purpose was to find out in how far inquiries of this kind are feasible from a psychological and technical point of view. The results of this supplementary survey have shown that in this way important information may be obtained for health policies which otherwise could not be secured. The morbidity data available in the Federal Republic originate for the most part from material not originally produced for use in statistics; they are rather provided by the public health offices from their records on notifiable diseases and by the legal health insurance on cases of incapacity for work. These data thus cover only the incidence of selected diseases

or the morbidity among part of the population, namely for the insured persons. The figures relate to cases rather than to diseased persons, so that statistical information on important demographic and occupational aspects for these persons is lacking.

15. Household inquiries, however, permit an insight into public health conditions and also account for the economic and social aspects which are important for cases of illness and accidents. Representative morbidity statistics covering the households may, accordingly, be a valuable supplementation of the data provided by the public health offices and the legal health insurance. These are the reasons why a first supplementary inquiry on "illness and accidents" was conducted 15 years ago within the framework of the current 1 per cent sample survey of the microcensus. About 125,000 households comprising approximately 340,000 persons were covered in the Federal Republic and asked whether "a member of the household had been ill, confined to bed or under medical treatment in April" or whether he/she had been "under permanent medical treatment" because of a chronic ailment or for surveillance of a previous disease. Another question was to provide information on whether "a member of the household had been involved in any accident whatsoever (e.g., traffic accident, industrial accident, household accident, etc.) or injured by poison during the period February to April". Information was collected from this group of persons, which had been delimitated by these guiding questions, on the type of the disease or accidental injury. It was further determined whether and how long the persons concerned had been under medical treatment, confined to bed, in hospital and/or incapable for work as a result of the disease or accident. For persons injured in accidents, information was collected on date, type, circumstances and consequences of the accident. Persons who had suffered in the report period from several diseases or had been involved in several accidents were counted only once with the most serious disease or accident. Questionnaires were not to be filled in for household members under permanent treatment by a lay medical practitioner. Routine cases of dental care and all negligible cases were also left out of consideration.

16. The brief comments given below on the definition of "illness" and the method employed are to facilitate the interpretation of the results. There is no clear and universal definition of "illness". Illness is mainly a subjective process. Whether a respondent is ill or not depends in the first place on his (her) own judgement. As, accordingly, illness cannot be covered in a household inquiry as an objective condition, the subjective opinion of the respondent had to provide the point of departure. The interviewers, who were not medically trained, had to take down the answers as they were given by the respondents, irrespective of any diverging opinions with regard to morbid conditions which objectively are of the same nature. So the results were also influenced by the subjective judgement of the respondents with regard to the state of health or illness of the household members. This influence decreases with the growing "gravity" of the disease or accident and the increasing number of objective consequences they have (confinement to bed, medical treatment, incapacity for work, hospitalization) in the individual instance.

17. Experience gained from similar inquiries on past events has shown that the information given by a respondent is influenced by two further factors, namely, by gaps of memory and the so-called telescoping effect. The former denotes that an event or the date of an event is the sooner forgotten, the greater the time distance between the event and the key-date of the inquiry becomes. The later

results in a shift of the date towards the present. Both effects may have biased in particular the figures on persons with accidental injuries, because they have been based on a reference period of three months.

18. In order to eliminate those factors, in 1970 the same reporting period for injured persons was chosen as for persons with diseases (one month), and the enquiry was not made on a previously decided random date, but on any day within the four weeks' survey period. The concept for 1972 was again modified so as to take the morbidity for the day of survey and/or the preceding four weeks' time interval. The concept for 1973 till 1980 followed this pattern too. The data, calculated accordingly, refer to the four weeks reporting period.

19. Answering the questions is voluntary. The collaboration of the population has been sufficiently good up to now. A maximum of 1.1 per cent of the population, however, made use of their right of refusal.

20. Although in 1966 it was possible to state more than one cause, only the most serious disease and/or injury was being further investigated. In 1970 acute and chronic diseases were separately enquired about. From 1972 onwards only the most serious disease was recorded, and whether it was chronic or acute. The types of diseases, injuries due to accidents, and disabilities are entered in full context, and then coded. The coding of the types of diseases took place in 1966 according to the two-digital system of the social insurance bodies, and from 1970 onwards according to a system based on the ICD, which was worked out specifically for health surveys. The following groups represent the nature of injuries due to accidents: bone fractures (including fracture of the skull), internal injuries (including damage of the brain), deep open wounds, burns and poisonings.

21. The definition of the cause of accident is a functional one. As no comparable international system is available, in the FRG the following categories have been defined:

1. Accident at work and on duty (excluding journey to and from work),
2. Accident at school or training (excluding journey to and from school),
3. Traffic accident or accident during journey to and from work/school,
4. Accident occurring at home,
5. Accident at sports/play grounds.

22. From 1976 onwards random samples for health surveys were no longer be taken as additional surveys, but as part of the original microcensus programme every two years (sampling fraction: 1 per cent). The Federal Republic Health Authority has recognized these surveys, and issued a statue for the following items to be included in a survey:

1. The most important risk factors and indicators of the population, e.g., environmental factors, pollution, smoking habits.
2. Use of the health services, e.g., frequency and purpose of consulting a doctor, and the benefit of preventive and early diagnostic methods.

23. The problems of health random samples can only briefly be indicated here. Many problems result on the one hand from putting questions to lay people, and on the other hand from carrying out the survey by lay people.

24. The decision whether or not a disease is present, is made by a lay person, and therefore depends on the wellbeing of the subject. There is little preparation for giving the requested information about certain diseases. This, up to now, had been the situation when asking questions to detect tuberculosis. Another source of incorrect answers is the fact that medical doctors do not always explain the illness clearly to their patients (e.g. malignant growths). The accuracy of statements is further reduced by expecting the interviewed person give information not only about his/her own disease/injury/disability, but also about those of the other household members.

25. The results of enquiries are recorded by medically untrained interviewers, another source of mistakes. Whether and how far the results are affected by the interviewers' technique of questioning, whether and how far their personal interest plays a role, requires further thorough investigation. Such surveys could possibly include random examinations by medical doctors. This would not only be a measure of accuracy but also of completeness of lay information.

26. Which diagnostic code to choose may be a matter of subjective judgement by the coders, and a further source of mistakes. The sampling fraction in 1970, 1972 and 1973 was 0.1 per cent. Therefore, considering a nationwide survey of such a large scale, many areas were only poorly covered and could not be evaluated because of the random fluctuations. This particularly refers to the relatively small number of persons injured in accidents, where the combined evaluation of age, sex, nature of injury and accident categories cannot be taken into account; nor is it possible in many cases to identify differential types of diseases in connexion with socio-economic data. Also the temporal comparison of certain survey characteristics is restricted hereby.

V. THE GERMAN HOUSING SAMPLE SURVEYS

27. The first 1 per cent - housing sample survey was conducted in the year 1957. Its main objective was to determine the extent and prevailing trends of the housing requirements by questionning households. Moreover, it served to reveal the relationships between income and rent, or income brackets and type of accommodation, respectively. Above and beyond the problems mentioned in the following sections, it was in the first years possible to clarify some methodological questions in the course of the housing survey. Thus, for example, it was to be established whether the technical definition of "dwelling" employed in the previous housing surveys in the Federal Republic of Germany had been correctly understood by the public: Households were asked whether they had a structurally planned (or conventional) kitchen, and from the answers to this question it was then decided for dwellings located in a permanent building whether the housing unit was on the one hand, a "conventional" dwelling or, on the other, an improvised dwelling. It was further necessary to check whether the definition of "household",

which is fundamental to the population census and to the sample surveys of official statistics, is also suitable for special housing investigations. Finally, it was to be determined whether and to what extent a sample survey in which dwellings are selected for the interviewer to visit can supply exact results valid over any length of time.

28. Another point in connexion with the definition of the dwelling is that it appeared that the public, in deciding whether a dwelling included a conventional or structurally planned kitchen or not let itself be influenced by a variety of considerations. The survey showed that in about 7 per cent of the cases in which households stated that their dwelling included a conventional kitchen, the kitchen had neither sink nor running water. As a result of this experience, it was necessary in future enquiries to see that the decision in individual cases as to whether a conventional dwelling is concerned or not is not made dependent on an opinion question. In the processing stage, a large number of characteristics obtained in answer to questions will have to be correlated on uniform principles to form types of housing units, in order to reach a classification of the entire housing stock according to installations and equipment, similar to that of the sample survey of 1960.

29. The definition of "household" has in fact proved suitable for enquiries into housing conditions. The principles for the delimitation of households applied in the Federal Republic of Germany entail from the start that only such groups of persons are considered to be households who live together and keep house together (e.g., live from a common budget). This is shown in the fact that housing wishes are relatively seldom entertained by parts of households, and that in cases in which the household is thinking of moving, a splitting up of the original household into a number of households is an exception. As far as subtenants are concerned, who form independent households, it must be remembered that, especially in rural circumstances, the accommodation of several households in what is structurally one housing unit often does not lead to unsanitary conditions and cannot in itself be regarded as undesirable. This is particularly true when occupier and subtenant are relatives. Hence, in many cases the delimitation of households may have been too rigidly carried through. In future surveys, more attention were paid to the economic and familial relations in these cases.

30. In the sample survey, using the results of the housing statistics of 1956, brought up to date by the help of current building statistics, structural housing units were pre-selected for the interviewer to visit. It has emerged, however, that in one-family and two-family houses in particular, the delimitation of structural housing units did not always conform to the actual principles recommended. It was also apparent that the interviewer in many cases had not only visited the prescribed housing units and interviewed all the households living in it, but had also included in the survey other dwellings and households in the respective building. Even if such errors are recognisable because of the nature of the survey lists, and can be eradicated, such experience leads to the opinion that in the following sample surveys, the whole building and all the households contained in it should be made object of the survey in each case. Such a survey technique would make it possible to supplement qualitative data on the dwellings, which are obtained in any case, by corresponding data on the building in which they are contained.

31. The 1972 and the 1978 housing sample survey were the first enumeration of its kind to include also a question programme relating to town planning. Here, official statistics followed the generally accepted knowledge that aspects of town planning are steadily gaining in importance as far as housing is concerned. The selection of a dwelling is particularly influenced by its surroundings, by emission and noise factors and by the availability of employment and public facilities. Statistically representative data on the location of the building, the use of real estate for building and other purposes and the relations of such data to social and economic conditions and negative environmental factors were not yet available. Normal individual investigations may well be suitable for the evaluation of local and regional conditions and requirements, but are of only limited value for gaining general insights. To obtain a reliable perspective, especially with respect to different kinds and degrees of developing real estate and their consequences for the population and housing, enquiries to that effect were indispensable.

32. In the 1978 sample survey on housing, linked with the microcensus, questions were put for the second time on the relationship of the dwellings to their environment, on noise (see Annex III, Question Nr. 9) or smell nuisances, etc. These are questions which taken in conjunction with the data on family structures are also significant for the procurement of social indicators; for example, questions on the subjective assessment of opportunities for children to play, shopping facilities, the distance from the place of work or the accessibility or otherwise of public transport (see Annex III, Question Nr. 10). In some cases these could be "checked" by objective questions, relating to the selected area concerned and submitted for answer to the administration of the relevant local area or town.

33. The 1 per cent housing sample surveys of 1972 and 1978 were taken in conjunction with the microcensus, which is a 1 per cent sample survey of the population. This had organizational, financial and evaluation technical advantages.

34. The selection was made from the material of the 1970 population census. With respect to the number of households to be interviewed, efforts were made to use the most homogeneous sampling units possible. The number of such households per sampling district was fixed at 20 households for communities below 20,000 inhabitants and at 30 households for larger communities.

35. Altogether, the communities included in the sample were allocated to eight community size classes and grouped in corresponding strata. Large buildings (with 25 households and above) and larger institutions (with 50 and more persons) constituted their own sampling units.

36. The sampling material arranged by communities, streets and house numbers, was formally subdivided into zones of 100 sampling units, of which one sampling unit was selected at random.

37. Unusual features were taken into account by an adaptation of this sampling principle.

VI. COMPARABILITY OF SOCIAL SURVEY DATA

38. As regards the comparability of social survey data, we can say:

- data comparability is assured only for the basic areas, with regard to the "classic" normative variables; questionnaire differences with regard to the precise description and disaggregations of data are generally substantial enough to make any comparative reading of the results a very difficult matter;
- as regards subjective and perceptive data, the differences between concepts, response coding and question formulation in different countries or different parts of a country make comparison of results practically impossible.

39. On the other hand, social surveys seem to be of considerable interest for the study of links between normative variables. Since the sample is fairly large in many cases and these surveys cover a considerable number of questions, it is possible to construct links between various indicator levels, which often cannot be done with other sources.

COMPREHENSIVE SURVEY (OECD): QUESTIONNAIRE ON HEALTH

Questionnaire

		COL/ CODE	ROUTE
HEALTH SECTION			
I would now like to ask you some questions about your health. Some of these questions may not be relevant to you, but it is necessary to ask a range of questions to include people with different levels of health.			
30 In general, how would you describe your state of health? Would you say it was:	very good	1	
	good	2	
	average	3	
	poor	4	
	very poor	5	
31 Do you suffer from any long-standing illness, effects of an injury, any disability or any weakness?	Yes	1	
	No	2	Q34
32 Could you describe it for me? NOTE THE CONDITIONS OR ILLNESSES MENTIONED AS CAREFULLY AS POSSIBLE Condition 1 ..... Condition 2 ..... Condition 3			
33 Do(es) the condition(s) you have cause your working capacity or your normal daily activities to be restricted?	Yes - to a great extent	1	
	Yes - to some extent	2	
	No - not at all	3	

		COL/ CODE	ROUTE
34	Could you run 100 metres? (As indicated) IF YES → Q36 Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	 } Q36
35	Can you walk 400 metres? (As indicated) Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	
36	Can you walk up and down stairs? Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	
37	Can you carry an object of 5 kilos for 10 metres? (As indicated) Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	
38	Can you cut your own food? (Such as meat, fruit) Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	
IF NO DIFFICULTY INDICATED IN Q35 to Q38 GO TO Q43 IF DIFFICULTY IS INDICATED ASK Q39 TO Q 42			
39	Can you, when standing, bend down and pick up a shoe from the floor? Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	
40	Can you dress and undress? Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	
41	Can you move between rooms? Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	

		COL/ CODE	ROUTE
42	Can you get in and out of bed? Yes, with no major difficulty Yes, but with difficulty No, not able to	 1 2 3	
43	Is your eyesight good enough to read ordinary newspaper print? (with glasses on if necessary) Yes Yes, with difficulty No	 1 2 3	
44	Is your eyesight good enough to see clearly the face of someone at the other side of the room? (approximately 4 metres) Yes Yes, with difficulty No	 1 2 3	
45	Can you hear what is said in normal conversation with one other person? (with hearing aid if necessary) Yes Yes, with difficulty No	 1 2 3	 Q47
46	Can you hear what is said in a normal conversation with at least 2 other persons? (with hearing aid if necessary) Yes Yes, with difficulty No	 1 2 3	
47	Can you both bite and chew on hard foods? (for example, a firm apple, celery) Yes Yes, with difficulty No	 1 2 3	

48 During the two weeks ending last Sunday have you had an illness or injury which:

a) caused you to stay in bed for at least part of one or more days?

(FILL IN BOX) If no, code 00
If yes, write in number of days

b) caused you to cut down on thing you usually do e.g., you did not go to work/school, did less housework, cancelled other

(FILL IN BOX)

	CCL/ CODE	ROUTE
If no, code 00 If yes, write in number of days		

NOTE: DO NOT INCLUDE DAYS IN BED IN 48 b)

Box for Q48

Last two weeks				
Cause	No. of days in bed	No. of days cut down activities		
Illness				
Injury				

49 During the last 12 months:

a) have you been admitted to a hospital or clinic because of illness or injury?

If no, code 0 in box
If yes, code number of days

b) have you been confined to bed because of illness or injury?

Do not include days in last two weeks Q48 a)

NOTE: DO NOT INCLUDE DAYS IN HOSPITAL/CLINIC IN Q49 b)

Box for Q49

Last 12 months				
Cause	Days in hospital/clinic	Days in bed		
Illness				
Injury by accident				
Injury by assault or attack				

	COL/ CODE	ROUTE
If respondent has spent time in hospital or bed in last 12 months because of injury by accident:		
50 Where did this accident take place?		
SHOW CARD B		
Injured – At home	1	
– At work	2	
– In traffic	3	
– In traffic at work	4	
– Playing sport (or exercising)	5	
– In public place (street, restaurant, pub, cinema, etc.)	6	
– Other (WRITE IN)		
.....		
.....		
.....		
NOTE: IF INJURED MORE THAN ONCE, INDICATE THE MOST SERIOUS INJURY		
If respondent has spent time in hospital or bed in last 12 months because of injury caused by assault or attack		
51 Where did this assault/attack take place?		
SHOW CARD C		
– At a work place	1	
– In a private home	2	
– In a restaurant/cafe/pub/bar	3	
– In a bus/train/metro/bus stop/station	4	
– In a private vehicle/taxi	5	
– Other public place/street/park	6	
– Other (WRITE IN)		
.....		
.....		
.....		
TO ALL RESPONDENTS		
52 During the past 12 months has someone assaulted or attacked you without you being injured		
Yes	1	
No	2	Q54
53 Where did this happen?		
SHOW SAME CARD AS Q51 (CARD C)		

		COL/ CODE	ROUTE
54	Are you afraid to take a walk in your neighbourhood at night?		
	Yes, very afraid	1	
	Yes, afraid	2	
	No	3	
55	Do you suffer from any condition (illness, disability, infirmity) which you have not mentioned in the above questions?		
	Yes	1	
	No	2	

If yes, please describe .....
.....
.....
.....
.....
.....

...l foundation
Covered by the Ordinance on Follow-up Progresses to the Microcensus dated 22 Dec. 1971 (BGBl. I, p. 2158)

Sicknesses and Accidents
Follow-up Survey to the Microcensus, April 1974, [illegible]
compliance is voluntary

Please use only pencil

Order characteristics ▶ Land | Actual district [illegible] | Interview Day Month | Questionnaire No. | 109002

	1st Person	2nd Person		1st person	2nd person
	ser. No.	Year of birth	male		
	ser. No.	Year of birth	female		

Only applying to persons in the households who were sick today or were sick during the last 4 weeks

4 Kind of illness (in the case of several diseases, state the most serious one)

1st person

2nd person

				1st person	2nd person
4	Duration of illness	1-3 days			
		4 days to 1 week			
		over 1 week to 2 weeks			
		over 2 weeks to 4 weeks			
		over 4 weeks to 6 weeks			
		over 6 weeks to 1 year			
		over 1 year			
5	Is the person still ill today?	Yes			
		No			
6	Is this illness a long-lasting and/or chronic affection (e.g. rheumatism, diabetes)?	Yes			
		No			

Only applying to persons in the household who were today or during the last 4 weeks hurt in an accident.

5 Kind of injury caused by an accident(in the case of several injuries state the most serious one)

1st person

2nd person

			1st person	2nd person
6a	Accident category			
	work accident (excl. accidents on the way to work			
	School accident (excl. accidents on the way to and back			
	Traffic accident (incl. accidents on the way to work/school and back			
	Domestic accident			
	Sport/playground accident			
	Other accident			
6b	If "work accident" or "domestic accident" in position 6a: was this accident related to house work	Yes		
		No		

7 Date of accident?

1st person Day Month | 2nd person Day Month

On account of illness/injuries caused by accident, during the last 4 weeks

			1st person	2nd person
8a	Stationary treatment in hospital? Yes, viz.	Illness		
		Accident		
		No		
8b	Today (day interview) still in hospital? Yes, viz.	Illness		
		Accident		
		No		
9	At home confined to bed (or was confined)? Yes, viz.	Illness		
		Accident		
		No		
10a	Was unable to work? Yes, viz.	Illness		
		Accident		
		No		
10b	Duration of incapability 1-3 days			
	over 3 days to 1 week			
	over 1 week to 2 weeks			
	over 2 weeks			
10c	Today (day of interview) still unable to work Yes, viz.	Illness		
		Accident		
		No		
11	Are you (or were you) receiving medical treatment other than stationary treatment in a hospital?	Illness		
		Accident		
		No		
12	If yes to question 11: by a general practitioner (a not specialised physician)?	Yes		
		No		
12b	by a specialist?	Yes		
		No		
12c	at the hospital (out-stations department or polyclinic)	Yes		
		No		

Handicapped persons in the household (Applies only handicapped persons born in 1957 and before! With regard to handicapped children born in 1958 and after, complete supplementary sheet "Handicaps of children".)

13 Kind of handicap

1st person

2nd person

			1st person	2nd person
14	Cause of handicap	Accident		
		War invalidity		
	Other causes (e.g. congenital defect)			
15	Reference person: Sick, injured or handicapped person his/her self	Yes		
		No		

TO BE COMPLETED BY THE [illegible] OF STATISTICAL OFFICE ONLY

Interview Day Month | Microcensus questionnaire No. | Substitute identification No.

ser. No. | Year of birth | Day Month | 1 | 5 | 13

ser. No. | Year of birth | Day Month | 1 | 5 | 13

Federal Statistical Office

Printed matter
No. 4

Please use
pencil only

HOUSING QUESTIONNAIRE

1 % - Housing Sample Survey 1978

1 If the dwelling is unoccupied: Is it in future to be used as a dwelling? — Yes — No —

▶ If so: Why is it unoccupied?
- New dwelling not yet occupied
- Because owner or tenant have changed
- Because of modernization, renovation or conversion

Is there already a definite future tenant or owner? — Yes — No —

Has the dwelling already been unoccupied for more than 3 months? — Yes — No —

A L L OCCUPANTS OF DWELLINGS

2 Is the dwelling
- a main dwelling
- a secondary dwelling (excl. questions 13 to 18)
- for purposes of studies/training
- for vocational/business purposes
- a leisure-time dwelling (Answer questions 3 to 6 only.)

3 How many living and bedrooms are there, incl. those outside the self-contained dwelling, but excl. kitchen?

Number of rooms: 1 2 3 4 5

- 6 to under 10 m²
- 10 to under 15 m²
- 15 to under 25 m²
- 25 m² and over
- Of which are sublet
- Moreover, commercially used rooms of the dwelling

4 What is the total size of the dwelling, incl. kitchen, bath, toilet, corridor, attic, chargeable balcony space and rooms sublet or commercially used?

(Size: m²)

5 Is the dwelling heated by
- Remote, block, central heating
- Multiple heating
- Single or multi-room stoves (also electric night storage heaters)

▶ If multiple heating, single or multi-room stoves: Kind of fuel?
- Coal, coke, etc.
- Fuel oil
- Gas
- Electricity

6 Has the dwelling
- Kitchen
- Kitchenette/cooking cabinet
- Only a cooking facility provisionally fitted in a room
- Bath/shower of its own
- Common bath/shower
- Flush toilet inside dwelling
- Flush toilet outside dwelling
- Toilet without flushing
- Hot-water supply, as central supply for building
- for dwelling
- Individual apparatus(es)
- Mainly double, compound or insulated-glass windows
- Balcony, loggia together 2 to under 6 m²
- 6 m² and over

(One — Several —)

7 How do you rate your dwelling? (S) = Subjective questions

a) Size:
- Sufficient
- Could be larger
- Is virtually too large

b) Equipment:
- Corresponds to the needs
- Should be better

c) Costs:
- Appropriate
- Rather high
- Low

8 a) How long does it take to walk to the nearest ...?

Walking minutes: up to 5 | 6 to 10 | 11 to 20 | more than 20 | Do not know

- Stop for public transport
- Shopping facilities
- Kindergarten
- Primary school
- Public children's playground
- General practitioner
- Greens open to the public (park/green areas/fields/meadows/woods)

b) How frequently do public means of transport depart from the nearest stop during the daytime? (Bus, train, tramway and the like, also several lines)
- Every few minutes
- Every 10 minutes
- Every 20 minutes
- About half-hourly
- About every hour
- Only several times a day
- Do not know

9 a) Is the dwelling located at a thoroughfare?
b) How is the motor traffic in your street during the daytime?
c) Is there in the direct vicinity ...

(S)

	Heavy	Less heavy	Little	Occasionally heavy	Occasionally light	None
Railway noise						
Road-traffic noise						
Air-traffic noise						
Industrial noise						
Smells/exhaust-gases/dust						

10 How do you rate your residential area as a whole with regard to:

(S)

	Good	Medium	Should be better	Inadequate	No opinion
Linking with public transport service					
Availability of green areas and open spaces					
Availability of public children's playgrounds					
Availability of shops (shopping facilities)					
Availability of primary schools					
Availability of kindergartens					
Medical care					
Traffic safety for children/elderly persons					
Health hazards due to noise					
Health hazards due to air pollution					

11 What would you consider the most important aspect of your present housing situation?
▶ If the dwelling is more important:

(S)

- The kind of housing area
- Your dwelling as such
- Is its location of hardly any importance to you
- or does it play a major role

12 Since when has the present head of household been living in this dwelling?

- Before 1965
- 1965 to 1971
- 1972 to 1975
- 1976
- 1977
- 1978

FOR OCCUPANTS OF MAIN DWELLINGS ONLY

13 Is this the first dwelling of your own? Yes — No —
▶ If so: How did the head of household live before?

- As mencer of a household
- As subtenant

14 Did you have to move when occupying this dwelling? Yes — No —
▶ If so: How did the head of household live before?

- In a rented dwelling
- In his/her own house, in his/her owner-occupied dwelling

-Questions 15 to 18 only if you have moved since 1972-

15 How often did the present head of household move since 1972?

Once — Twice — Three times — More than three times —

16 Was the last move of the household due to a change of place of work/training of the main breadwinner? Yes — No —
▶ If so: How does the new place of work compare with the previous one?

	Better	Same	Worse
Way to work			
Security of the place of work			
Income			
Amenity of work			
Growth possibilities			

17 a) Was the last move due to a rent increase or a notice to leave concerning the previous dwelling of the head of household? Rent increase — Notice to leave —
▶ If notice to leave: Was the building to be pulled down? Yes — No —
b) Was the move due to a change of size of the household? Enlargement — Diminution —
c) What other reasons for the move were involved?

	No	1st reason	2nd reason
Desire to own the dwelling			
Size of dwelling inadequate			
Equipment of dwelling inadequate			
Dwelling too expensive as compared with the income			
Unfavourable location of the dwelling with regard to place of work			
with regard to training facilities			
with regard to shopping facilities			
Noise, exhaust-gases, dirt			
Not enough green areas			
Happened to get dwelling/house under favourable circumstances			
Other reasons			

Annex IV

THE 1980 OECD LIST OF SOCIAL INDICATORS

SOCIAL CONCERN		INDICATOR
HEALTH		
Length of life	1	Life Expectancy at age 1, (20,40,60)
	2	Perinatal mortality rate
Healthfulness of life	3	Short-term disability
	4	Long-term disability
EDUCATION AND LEARNING		
Knowledge, skills, values necessary to function in society	5	Literacy rate
The [illegible] of organised educational opportunities	6	Years of regular education
	7	Participation in adult education
EMPLOYMENT AND QUALITY OF WORKING LIFE		
Availability of gainful employment for those who desire it	8	Unemployment rate
	9	Rate of involuntary part-time work
	10	Rate of voluntary part-time work
	11	Discouraged workers rate
Quality of working life	12	Fatal work injury rate
	13	Non-fatal work injury rate
	14	Physical work environment nuisances
	15	Excessive work-related physical stress
Work related time	16	Average working hours
	17	Travel time to work
	18	Shift work and unsocial hours
	19	Paid annual leave
TIME AND LEISURE	20	Free time
The use of time	21	Free time activities
COMMAND OVER GOODS AND SERVICES		
Distribution of income	22	Dispersion of income
	23	Final income shares
	24	Average earnings
Distribution of wealth	25	Dispersion of wealth
Material deprivation	26	Low incomes
	27	Share of population below the poverty line
PHYSICAL ENVIRONMENT		
The man-made environment Housing conditions	28	Indoor dwelling space
	29	Access to outdoor space
	30	Households without basic amenities
Physical accessibility to services	31	Proximity of selected services
Environmental nuisances	32	Exposure to air pollutants
	33	Exposure to noise
SOCIAL ENVIRONMENT		
Social attachment	34	Family situation of children
	35	Suicide rate
PERSONAL SAFETY AND THE ADMINISTRATION OF JUSTICE		
Victimisation	36	Deaths from unexpected events
	37	Serious injuries from unexpected events
	38	Fear for personal safety
SOCIAL OPPORTUNITY AND PARTICIPATION		
Social mobility	39	Socio-economic mobility rate